Simple Correlation and Regression analysis

Simple Correlation and Regression Analysis

→ The relationship between the two variables such that change in the value of one variable, makes a change in the value of the other variable is known as the correlation.

→ The measure of the degree of correlation between the two variables is known as the correlation coefficient.

* Types of correlation:
  1. Positive and Negative correlation
  2. Linear and non-linear correlation
  3. Simple, multiple and partial correlation.

* Methods of studying correlation:
  1. Scatter diagram
  2. Karl Pearson's correlation coefficient
  3. Spearman's Rank correlation

* Karl Pearson's Correlation coefficient:

It is also known as Pearson's correlation coefficient.

$$r = \frac{\mathrm{cov}(X.Y)}{\sqrt{\mathrm{var}\,X}\ \sqrt{\mathrm{Var.}\,y}}$$

Note that r is also called pearson's coefficient of correlation.

Properties of correlation coefficient:

1. The value of r has no unit.

2. Its formula is symmetrical: $r_{xy} = r_{yx}$

3. It is independent of the change of origin: $r_{xy} = r_{UV}$

where.

$$U = X - A$$

$$V = Y - B \quad \text{and } A.B = \text{Assumed means}$$

4. It is independent of the change of scale: $r_{xy} = r_{uv}$

where.

$$u = \frac{X-A}{h}$$

$$v = \frac{Y-B}{K} \quad \text{and } h, k = \text{constants}$$

5. Its values lies between -1 to 1 i.e $-1 \le r \le 1$

6. If $r = 1$, there is perfect positive relationship
   If $r = -1$, there is perfect negative relationship
   if $r = 0$, there is no correlation at all.

7. It is the geometric mean between two regression coefficient i.e $r = \sqrt{b_{xy} \cdot b_{yx}}$

Computing coefficient of correlation:

(i) When deviations from actual mean are used:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}}$$

(ii) When actual data are used: Direct method

$$r = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{\left[n\Sigma X^2 - (\Sigma X)^2\right]\left[n\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

OR

$$r = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\;\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

(iii) When deviations from assumed mean are used

$$r = \frac{n\Sigma UV - \Sigma U \cdot \Sigma V}{\sqrt{n\Sigma U^2 - (\Sigma U)^2}\ \sqrt{n\Sigma V^2 - (\Sigma V)^2}}$$

Where.

$U = X - A$

$V = Y - B$

A, B = Assumed means

(iv) When step-deviations are used

$$r = \frac{n\Sigma uv - \Sigma u \cdot \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\ \sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

Where.

$u = \frac{X - A}{h}$

$v = \frac{Y - B}{K}$

h, K = Common factors

* Probable Error

Probable error of the correlation coefficient denoted by P.E. is the measure of testing the reliability of the calculated value of r.

$$P.E = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

Note:

i. If $r < P.E$, it is insignificant.

ii. If $r > 6P.E$, it is significant.

* Correlation coefficient of bi-variate distribution.

→ The distribution in which the values of two variables X and Y are grouped and the frequencies of different groups are given is known as "Bivariate frequency table" or "Correlation table".

→ The correlation coefficient for bivariate distribution is computed by the following formula:

$$r = \frac{N\Sigma fuv - \Sigma fu \cdot \Sigma fv}{\sqrt{N\Sigma fu^2 - (\Sigma fu)^2}\ \sqrt{N\Sigma fv^2 - (\Sigma fv)^2}}$$

where,

N = Total frequency, $u = \frac{X-A}{h}$, $V = \frac{Y-B}{K}$

* Rank Correlation:

The degree of relationship between two variables with respect to their respective ranks is known as "Rank Correlation coefficient"

$$R = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$$

Where,

$d = R_1 - R_2$

n = no. of pair of observations

Note: The value of R lies between -1 and $+1$.

Repeated Ranks

$$R = 1 - \frac{6\left[\Sigma d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \cdots\cdots\right]}{n(n^2-1)}$$

* Regression Analysis

- It is a statistical device, with the help of which we can estimate or predict the value of one variable when the value of other variable is known.

- The unknown variable which we have to predict is called dependent variable and the variable whose value is known is called independent variable.

→ The analysis used to describe the average relationship between two variables is known as simple linear regression analysis.

- Regression of X on Y
- Regression. of Y on X

Regression equation and regression coefficient:

a) The regression equation of Y on X which is used to describe the variation in the value of Y for given change in the value of X.

Equation:

$$Y - \bar{Y} = b_{yx}(x - \bar{x})$$

where,

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \cdot \Sigma Y}{n\Sigma x^2 - (\Sigma x)^2}$$

OR,

$$b_{yx} = \frac{n\Sigma uv - \Sigma u \cdot \Sigma V}{n\Sigma u^2 - (\Sigma u)^2}$$

$$b_{yx} = \frac{N\Sigma fuv - \Sigma fu \cdot \Sigma fv}{N\Sigma fu^2 - (\Sigma fu)^2} \times \frac{k}{h}$$ (Bivariate Distribution)

(b) The regression equation of X on Y which is used to describe the variation in the value of X for given change in the value of Y.

Equation:

$$X - \bar{X} = b_{xy} (Y - \bar{Y})$$

where,

$$b_{xy} = \frac{n \Sigma xy - \Sigma x \cdot \Sigma y}{N \Sigma y^2 - (\Sigma y)^2}$$

or. $$b_{xy} = \frac{n \Sigma uv - \Sigma u \cdot \Sigma v}{n \Sigma v^2 - (\Sigma v)^2}$$

or. $$b_{xy} = \frac{N \Sigma fuv - \Sigma fu \cdot \Sigma fv}{N \Sigma fv^2 - (\Sigma fv)^2} \times \frac{h}{k}$$ (Bivariate distribution)

Note:

The correlation coefficient is the square root of the product of the two regression coefficient i.e

$$r = \sqrt{b_{yx} \cdot b_{xy}}$$

2072 old (II) Q.No.8

Soln

Let. X = Age of husbands

Y = Age of wife

Calculation of Regression

Age of Husbands			20-30	30-40	40-50	50-60	60-70	$U = \frac{X-45}{10}$, $V = \frac{Y-40}{10}$			
Mid-value (X)			25	35	45=A	55	65				
Age of Wives	Mid-value (Y)	U / V	-2	-1	0	1	2	f	fV	fV^2	fUV
15-25	20	-2	$\boxed{20}$ 5	$\boxed{18}$ 9	$\boxed{0}$ 3	-	-	17	-34	68	38
25-35	30	-1	-	$\boxed{10}$ 10	$\boxed{0}$ 25	$\boxed{-2}$ 2	-	37	-37	37	8
35-45	B=40	0	-	$\boxed{0}$ 1	$\boxed{0}$ 12	$\boxed{0}$ 2	-	15	0	0	0
45-55	50	1	-	-	$\boxed{0}$ 4	$\boxed{16}$ 16	$\boxed{10}$ 5	25	25	25	26
55-65	60	2	-	-	-	$\boxed{8}$ 4	$\boxed{8}$ 2	6	12	24	16
		f	5	20	44	24	7	N = 100	$\Sigma fV = -34$	$\Sigma fV^2 = 154$	$\Sigma fUV = 88$
		fU	-10	-20	0	24	14	$\Sigma fU = 8$			
		fU^2	20	20	0	24	28	$\Sigma fU^2 = 92$			
		fUV	20	28	0	22	18	$\Sigma fUV = 88$			

Estimation of Age of wife.

$X = 75$ Years. $Y = ?$

The regression equation of Y on X is given by:

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 36.6 = 0.993(X - 45.8)$$

$$Y = 0.993X - 45.4794 + 36.6$$

$\therefore$ $Y = -8.8794 + 0.993X$ is the required Equation.

Working Notes:

$$\bar{X} = A + \frac{\Sigma fU}{N} \times h = 45 + \frac{8}{100} \times 10 = 45.8$$

$$\bar{Y} = B + \frac{\Sigma fV}{N} \times k = 40 + \frac{(-34)}{100} \times 10 = 36.6$$

$$b_{yx} = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{N\Sigma fU^2 - (\Sigma fU)^2} \times \frac{k}{h}$$

$$= \frac{100 \times 88 - 8 \times (-34)}{100 \times 92 - (8)^2} \times \frac{10}{10} = 0.993$$

Now,

The age of wife (Y) $= -8.8794 + 0.993 \times 75$

$= 65.5956$ Years

2069 Q.N.8

Sol^n

Let. X = height of father

Y = height of Son

$U = \frac{X-65}{2}$, $V = \frac{Y-67}{2}$

Height of sons			62-64	64-66	66-68	68-70	70-72				
Mid-value (Y)			63	65	B=67	69	71				
Height of Father	Mid-value (X)	U \ V	-2	-1	0	1	2	f	fu	fu^2	fuv
60-62	61	-2	$1\ \boxed{4}$	$2\ \boxed{4}$	-	$1\ \boxed{-2}$	$3\ \boxed{-12}$	7	-14	28	-6
62-64	63	-1	-	$1\ \boxed{1}$	-	$2\ \boxed{-2}$	$1\ \boxed{-2}$	4	-4	4	-3
64-66	A=65	0	$2\ \boxed{0}$	$3\ \boxed{0}$	$2\ \boxed{0}$	$1\ \boxed{0}$	-	8	0	0	0
66-68	67	1	-	$1\ \boxed{-1}$	$1\ \boxed{0}$	-	$1\ \boxed{2}$	3	3	3	1
68-70	69	2	$1\ \boxed{-4}$	$2\ \boxed{-4}$	-	-	-	3	6	12	-8
		f	4	9	3	4	5	N=25	$\Sigma fU = -9$	$\Sigma fU^2 = 47$	$\Sigma fUV = -16$
		fV	-8	-9	0	4	10	fV = -3			
		fV^2	16	9	0	4	20	fV^2 = 49			
		fUV	0	0	0	-4	-12	fUV = -16			

The regression coefficient of Y on X is

$$b_{yx} = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{N\Sigma fU^2 - (\Sigma fU)^2} \times \frac{k}{h}$$

$$= \frac{25 \times (-16) - (-9) \times (-3)}{25 \times 47 - (-9)^2} \times \frac{2}{2}$$

$$= -0.390$$

The regression coefficient of X on Y is

$$b_{xy} = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{N\Sigma fV^2 - (\Sigma fV)^2} \times \frac{h}{k}$$

$$= \frac{25 \times (-16) - (-9) \times (-3)}{25 \times 49 - (-3)^2} \times \frac{2}{2}$$

$$= -0.351$$

The correlation coefficient between X and Y is

$$r = \pm \sqrt{b_{xy} \times b_{yx}}$$

$$= \pm \sqrt{(-0.351) \times (-0.390)}$$

$$= \pm 0.370$$

$$\therefore r = -0.370$$

$\therefore$ There is low degree of negative Correlation between the height of fathers and sons.

2076 Q.No.18

Sol^n

Let. X = Motor Registration in (000 nos.)
Y = No. of tyres sold in (000 nos.)

Computation Table

Years	X	Y	XY	X^2	Y^2
1	60	68	4080	3600	4624
2	62	60	3720	3844	3600
3	65	62	4030	4225	3844
4	70	80	5600	4900	6400
5	48	40	1920	2304	1600
6	53	52	2756	2809	2704
7	73	62	4526	5329	3844
8	65	60	3900	4225	3600
9	82	81	6642	6724	6561
10	72	85	6120	5184	7225
N = 10	ΣX = 650	ΣY = 650	ΣXY = 43,294	ΣX^2 = 43,144	ΣY^2 = 44002

1. Calculation of coefficient of correlation between X and Y.

$$r = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{10 \times 43{,}294 - 650 \times 650}{\sqrt{10 \times 43{,}144 - (650)^2}\ \sqrt{10 \times 44002 - (650)^2}}$$

$$= \frac{432940 - 422500}{94.5515 \times 132.363}$$

$$= 0.83419$$

② Since, $r = 0.83419$, there is high degree of positive correlation between motor registration and no. of tyres sold.

③ Test for significant

$$PE(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1-(0.83419)^2}{\sqrt{10}}$$

$$= 0.0649$$

$$6PE(r) = 6 \times 0.0649 = 0.3894$$

Since: $r = 0.83419 > 6PE(r) = 0.3894$, r is significant.

④ calculation of two regression coefficient

Regression coefficient of X on Y is

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2} = \frac{10 \times 43294 - 650 \times 650}{10 \times 44002 - (650)^2} = 0.5959$$

Regression coefficient of Y on X is

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{10 \times 43294 - 650 \times 650}{10 \times 43144 - (650)^2}$$

$$= 1.1678$$

(5) Expected motor Registration = 92,000 (X)
Number of tyres (Y) = ?

The regression equation of Y on X is

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 65 = 1.1678(X - 65)$$

$$Y - 65 = 1.1678X - 75.907$$

$Y = -10.907 + 1.1678X$ is the required equation

Working Notes:

$$\bar{X} = \frac{\Sigma X}{N} = \frac{650}{10} = 65$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{650}{10} = 65$$

Now,

Number of tyres $(Y) = -10.907 + 1.1678 \times 92(000)$

$$= -10.907 + 107.4376\ (000)$$

$$= 96,530.6$$

2075 Q.NO.18

Soln

Let. X = Sales Revenue (000 Rs.)

Y = Advertisement expenditures (000 Rs.)

$U = \frac{X-150}{50}$, $V = \frac{Y-20}{10}$

Advertisement Exp.			5-15	15-25	25-35	35-45				
Mid-value (Y)			10	B=20	30	40				
Sales Revenue	Mid-value (X)	U \ V	-1	0	1	2	f	fU	fU^2	fUV
75-125	100	-1	4 $\boxed{4}$	1 $\boxed{0}$	-	-	5	-5	5	4
125-175	A=150	0	7 $\boxed{0}$	6 $\boxed{0}$	2 $\boxed{0}$	1 $\boxed{0}$	16	0	0	0
175-225	200	1	1 $\boxed{-1}$	3 $\boxed{0}$	4 $\boxed{4}$	2 $\boxed{4}$	10	10	10	7
225-275	250	2	1 $\boxed{-2}$	1 $\boxed{0}$	3 $\boxed{6}$	4 $\boxed{16}$	9	18	36	20
		f	13	11	9	7	N=40	$\Sigma fU = 23$	$\Sigma fU^2 = 51$	$\Sigma fUV = 31$
		fV	-13	0	9	14	$\Sigma fV = 10$			
		fV^2	13	0	9	28	$\Sigma fV^2 = 50$			
		fUV	1	0	10	20	$\Sigma fUV = 31$			

The coefficient of correlation between X and Y is

$$r = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{\sqrt{N\Sigma fU^2 - (\Sigma fU)^2}\ \sqrt{N\Sigma fV^2 - (\Sigma fV)^2}}$$

$$= \frac{40 \times 31 - 23 \times 10}{\sqrt{40 \times 51 - (23)^2}\ \sqrt{40 \times 50 - (10)^2}}$$

$$= \frac{1240 - 230}{38.871 \times 43.588} = 0.5961$$

∴ There is a positive correlation between sales revenue and advertisement expenditure.

Test for significance:

$$PE(r) = 0.6745 \times \frac{1 - r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1 - (0.5961)^2}{\sqrt{40}}$$

$$= 0.06875$$

Now,

$$6\ PE(r) = 6 \times 0.06875 = 0.4125$$

Since, $r = 0.5961 > 6PE(r) = 0.4125$, it is significant.

Estimation of sales Revenue:

Advertisement cost (Y) = Rs. 70 (thousand)

Sales Revenue (X) = ?

The regression equation of X on Y is

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

$$X - 178.75 = 2.658(Y - 22.5)$$

$$X - 178.75 = 2.658Y - 59.805$$

$$X = 178.75 - 59.805 + 2.658Y$$

$\therefore X = 118.945 + 2.658Y$ is the required equation.

Working Note:

$$\bar{X} = A + \frac{\Sigma fU}{N} \times h = 150 + \frac{23}{40} \times 50 = 178.75$$

$$\bar{Y} = B + \frac{\Sigma fV}{N} \times k = 20 + \frac{10}{40} \times 10 = 22.5$$

$$b_{xy} = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{N\Sigma fV^2 - (\Sigma fV)^2} \times \frac{h}{k} = \frac{40 \times 31 - 23 \times 10}{40 \times 50 - (10)^2} \times \frac{50}{10}$$

$$= 2.658$$

Now

Sales Revenue (X) = 118.945 + 2.658 × 70

= 305.005 (thousands)

= 305.005 × 1000

= Rs. 305,005

2074 Q.NO.19

Soln

Let, X = Advertisement Expenditure

Y = Sales

$U = \frac{X-45}{10}$, $V = \frac{Y-17.5}{5}$

Advertisement Exp.			20-30	30-40	40-50	50-60	60-70				
Mid value (X)			25	35	A= 45	55	65				
Sales	Mid-value (Y)	V \ U	-2	-1	0	1	2	f	fV	fV^2	fUV
10-15	12.5	-1	–	–	–	3 $\boxed{-3}$	7 $\boxed{-14}$	10	-10	10	-17
15-20	B= 17.5	0	–	4 $\boxed{0}$	9 $\boxed{0}$	4 $\boxed{0}$	3 $\boxed{0}$	20	0	0	0
20-25	22.5	1	7 $\boxed{-14}$	6 $\boxed{-6}$	12 $\boxed{0}$	5 $\boxed{5}$	–	30	30	30	-15
25-30	27.5	2	3 $\boxed{-12}$	10 $\boxed{-20}$	19 $\boxed{0}$	8 $\boxed{16}$	–	40	80	160	-16
		f	10	20	40	20	10	N = 100	$\Sigma fV = 100$	$\Sigma fV^2 = 200$	$\Sigma fUV = -48$
		fU	-20	-20	0	20	20	$\Sigma fU = 0$			
		fU^2	40	20	0	20	40	$\Sigma fU^2 = 120$			
		fUV	-26	-26	0	18	-14	$\Sigma fUV = -48$			

Correlation between advertisement expenditure and sales:

$$r = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{\sqrt{N\Sigma fU^2 - (\Sigma fU)^2}\ \sqrt{N\Sigma fV^2 - (\Sigma fV)^2}}$$

$$= \frac{100 \times (-48) - 0 \times 100}{\sqrt{100 \times 120 - (0)^2}\ \sqrt{100 \times 200 - (100)^2}}$$

$$= \frac{-4800}{109.55 \times 100}$$

$$= -0.4381$$

Test for significant:

$$P.E.(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1-(-0.4381)^2}{\sqrt{100}}$$

$$= 0.05450$$

$$6\ PE(r) = 6 \times 0.05450 = 0.3270$$

Since, $|r| = 0.4381 > 6\ PE(r) = 0.3270$, it is significant.

Estimation of Sales:

Advertisement Expenditure (x) = 82 crores

Sales (Y) = ?

The regression equation Y on X is

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 22.5 = -0.2(X - 45)$$

$$Y = 22.5 - 0.2X + 9$$

$\therefore Y = 31.5 - 0.2X$ is the required equation.

Working Notes:

$$\bar{X} = A + \frac{\Sigma fU}{N} \times h = 45 + \frac{0}{100} \times 10 = 45$$

$$\bar{Y} = B + \frac{\Sigma fV}{N} \times K = 17.5 + \frac{100}{100} \times 5 = 22.5$$

$$b_{yx} = \frac{N\Sigma fUV - \Sigma fU \cdot \Sigma fV}{N\Sigma fU^2 - (\Sigma fU)^2} \times \frac{k}{h}$$

$$= \frac{100 \times (-48) - 0 \times 100}{100 \times 120 - (0)^2} \times \frac{5}{10} = -0.2$$

Again,

Sales (Y) = $31.5 - 0.2 \times 82 = 15.1$ crores.

2072 (ii) Q. No. 19

Soln. Let.

X = Frequency of advertisement in electronic media/day

Y = Volume of sales per day.

Frequency of Adv.(X)			0	2	A = 4	6	8	$U = \frac{X-4}{2}$, $V = \frac{Y-7.5}{5}$			
Sales	Mid-val. (Y)	V \ U	−2	−1	0	1	2	f	fV	fV^2	fUV
0-5	2.5	−1	$\boxed{4}$ 2	–	–	–	–	2	−2	2	4
5-10	B = 7.5	0	–	$\boxed{0}$ 4	$\boxed{0}$ 5	$\boxed{0}$ 3	$\boxed{0}$ 4	16	0	0	0
10-15	12.5	1	–	–	–	$\boxed{4}$ 4	$\boxed{12}$ 6	10	10	10	16
15-20	17.5	2	–	–	–	–	$\boxed{8}$ 2	2	4	8	8
		f	2	4	5	7	12	N = 30	Σfv = 12	Σfv^2 = 20	ΣfUV = 28
		fU	−4	−4	0	7	24	ΣfU = 23			
		fU^2	8	4	0	7	48	ΣfU^2 = 67			
		fUV	4	0	0	4	20	ΣfUV = 28			

Since, $r = 0.6863 > 6\,PE(r)$, the relationship is significant.

b. Since, the relationship between advertisement on electronic media and sales is positive, frequency of advertisement on electronic media should be increased to promote the sales.

c. Estimation of sales:

Frequency of Advertisement $(X) = 7$

Sales $(Y) = ?$

The regression equation Y on X is given by

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 9.5 = 0.952(X - 5.53)$$

$$Y = 9.5 + 0.952X - 5.26456$$

$$Y = 4.23544 + 0.952X$$ is the required equation.

Working Notes:

$$\bar{X} = A + \frac{\Sigma fu}{N} \times h$$

$$= 4 + \frac{23}{30} \times 2$$

$$= 5.53$$

$$\bar{Y} = B + \frac{\sum fv}{N} \times k$$

$$= 7.5 + \frac{12}{30} \times 5$$

$$= 9.5$$

$$b_{yx} = \frac{N\sum fuv - \sum fu . \sum fv}{N\sum fu^2 - (\sum fu)^2} \times \frac{k}{h}$$

$$= \frac{30 \times 28 - 23 \times 12}{30 \times 67 - (23)^2} \times \frac{5}{2}$$

$$= 0.952$$

Now,

Sales $(Y) = 4.23544 + 0.952 \times 7$

$= 10.89944$ (in 000)

2068 Q. No. 11

Soln

Given:

$N = 25$, $\sum X = 125$, $\sum Y = 100$, $\sum X^2 = 650$, $\sum Y^2 = 460$, $\sum XY = 508$

Wrong pairs of observations : $(X, Y) = (6, 14)$ and $(8, 6)$

Correct pairs of observations : $(X, Y) = (8, 12)$ and $(6, 8)$

a. Calculation of correct values:

Correct $N = 25$

Correct $\Sigma X = 125 - 6 - 8 + 8 + 6 = 125$

Correct $\Sigma Y = 200 - 14 - 6 + 12 + 8 = 100$

Correct $\Sigma X^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$

Correct $\Sigma Y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436$

Correct $\Sigma XY = 508 - (6 \times 14) - (8 \times 6) + (8 \times 12) + (6 \times 8) = 520$

b. Calculation of correct coefficient of correlation:

$$r = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{25 \times 520 - 125 \times 100}{\sqrt{25 \times 650 - (125)^2}\sqrt{25 \times 436 - (100)^2}}$$

$$= \frac{13000 - 12500}{25 \times 30}$$

$$= 0.667$$

c. Calculation of equation of two lines of regression:

$$\bar{X} = \frac{\Sigma X}{N} = \frac{125}{25} = 5$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{100}{25} = 4$$

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2} = \frac{25 \times 520 - 125 \times 100}{25 \times 436 - (100)^2}$$

$$= 0.556$$

$$b_{yx} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{25 \times 520 - 125 \times 100}{25 \times 650 - (125)^2} = 0.8$$

The regression equation X on Y is

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

$$X - 5 = 0.556(Y - 4)$$

$$X - 5 = 0.556Y - 2.224$$

$$X = 5 + 0.556Y - 2.224$$

$$\boxed{\therefore X = 2.776 + 0.556Y}$$

The regression equation Y on X is

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 4 = 0.8(X - 5)$$

$$Y - 4 = 0.8X - 4$$

$$Y = 4 + 0.8 - 4$$

$$\boxed{\therefore Y = 0.8X}$$

d. Calculation of probable error:

$$P.E(r) = 0.6745 \times \frac{1 - r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1 - 0.667^2}{\sqrt{25}}$$

$$= 0.075$$

$$6\ P.E(r) = 6 \times 0.075 = 0.45$$

Since, $r = 0.667 > 6\ P.E(r) = 0.45$, the relationship is significant.

2073 Q.NO.18

Soln

Let. X = Job performance Index

Y = Salary

Computation Table

X	Y	X^2	Y^2	XY
9	36	81	1296	324
7	25	49	625	175
8	33	64	1089	264
4	15	16	225	60
7	28	49	784	196
5	19	25	361	95
5	20	25	400	100
6	22	36	484	132
$\Sigma X = 51$	$\Sigma Y = 198$	$\Sigma X^2 = 345$	$\Sigma Y^2 = 5264$	$\Sigma XY = 1346$

Employees Job performance (x) = 10 and 2

Salary (Y) = ?

The regression equation Y on x is given by:

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y - 24.75 = 4.2138(X - 6.375)$$

$$Y = 24.75 + 4.2138X - 26.863$$

$Y = -2.113 + 4.2138X$ is the required equation

Working Notes:

$$\bar{X} = \frac{\Sigma X}{N} = \frac{51}{8} = 6.375$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{198}{8} = 24.75$$

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{8 \times 1346 - 51 \times 198}{8 \times 345 - (51)^2}$$

$$= \frac{10768 - 10098}{2760 - 2601}$$

$$= 4.2138$$

Estimation of Salary:

If Job performance Index (X) = 10

Salary (Y) = $-2.113 + 4.2138 \times 10$

$= 40.025$ (000)

If Job performance Index (X) = 2

Salary (Y) = $-2.113 + 4.2138 \times 2$

$= 6.3146$ (000)

Test For significant:

The relationship between job performance Index and salary is:

$$r = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{8 \times 1346 - 51 \times 198}{\sqrt{8 \times 345 - (51)^2}\ \sqrt{8 \times 5264 - (198)^2}}$$

$$= \frac{10768 - 10098}{12.6095 \times 53.9259}$$

$$= 0.9853$$

$$P.E(r) = 0.6745 \times \frac{1 - r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1 - (0.9853)^2}{\sqrt{8}}$$

$$= 0.0069$$

$$6\,PE(r) = 6 \times 0.0069 = 0.0414$$

Since. $r = 0.9853 > 6 \cdot PE(r) = 0.0414$, the relationship is Significant.

2072 Q.NO. 19

Soln

Let. X = Promotion expenses

Y = Sales

Computation Table

Year	Y	X	X^2	Y^2	XY
2003	16	4	16	256	64
2004	20	4	16	400	80
2005	18	6	36	324	108
2006	24	10	100	576	240
2007	20	10	100	400	200
2008	22	12	144	484	264
N = 6	$\Sigma Y = 120$	$\Sigma X = 46$	$\Sigma X^2 = 412$	$\Sigma Y^2 = 2440$	$\Sigma XY = 956$

a. Calculation of two regression coefficient:

(i) Y on X

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2} = \frac{6 \times 956 - 46 \times 120}{6 \times 412 - (46)^2}$$

$$= 0.6067$$

(ii) X on Y

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2} = \frac{6 \times 956 - 46 \times 120}{6 \times 2440 - (120)^2}$$

$$= 0.90$$

b. Calculation of correlation coefficient between sales and promotional expenditure

$$r = \pm \sqrt{b_{yx} \cdot b_{xy}}$$

$$= \pm \sqrt{0.6067 \times 0.90}$$

$$= +0.7389$$

Since, $r = 0.7389$, there is high degree of positive correlation between sales and promotional expenditure.

c. Test for significance:

$$P.E(r) = 0.6745 \times \frac{1 - r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1 - (0.7389)^2}{\sqrt{6}}$$

$$= 0.1250$$

$$6\,P.E(r) = 6 \times 0.1250 = 0.75$$

Since, r is nearly equal to 6 PE(r), nothing can be concluded.

d. Estimation of Sales:

Promotional expenses (X) = Rs.20,000

Sales (Y) = ?

The regression equation Y on X is

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$
$$Y - 20 = 0.6067(X - 7.667)$$
$$Y = 20 + 0.6067X - 4.652$$
$$Y = 15.348 + 0.6067X$$ is the required equation.

Working Note:

$$\bar{X} = \frac{\Sigma X}{N} = \frac{46}{6} = 7.667$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{120}{6} = 20$$

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{6 \times 956 - 46 \times 120}{6 \times 412 - (46)^2}$$

$$= 0.6067$$

Now,

$$\text{Sales}(Y) = 15.348 + 0.6067 \times 20(\underline{000})$$

$$= 27.482\ (000)$$

e. The equation $Y = 15.348 + 0.6067X$ is in the form of $Y = a + bX$.

$a = 15.348$ indicates the sales when the promotional expenses is zero.

$b_{yx} = 0.6067$ indicates the rate of change of sales when the unit change in the promotional expenses. That is if the promotional expenses is increased by Rs. 1000, the sales is increased by Rs. 60,670.